
3C

Action

Professional Services Committee

Proposed Approval of an Alternative Teaching Performance Assessment, the Performance Assessment for California Teachers (PACT)

Executive Summary: This agenda item presents a recommendation for approval of an alternative teaching performance assessment, the Performance Assessment for California Teachers (PACT), in accordance with the Commission's responsibilities under Education Code Section 44320.2 (d)(3).

Recommended Action: Staff recommends the Commission approve the Performance Assessment for California Teachers (PACT) as an alternative teaching performance assessment.

Presenter: Phyllis Jacobson, Ed.D.,
Administrator, Professional Services Division

Strategic Plan Goal: 1

Promote educational excellence through the preparation and certification of professional educators.

- ◆ Sustain high quality standards for the preparation and performance of professional educators and for the accreditation of credential programs.

October 2007

Proposed Approval of an Alternative Teaching Performance Assessment, the Performance Assessment for California Teachers (PACT)

Introduction

Education Code Section 44320.2 specifies that all teacher preparation programs must implement a teaching performance assessment approved by the Commission commencing July 1, 2008. For purposes of carrying out this requirement, the statute allows institutions to either (a) implement the Commission's teaching performance assessment; (b) voluntarily develop an alternative teaching performance assessment for approval by the Commission; or (c) implement an alternative teaching performance assessment developed by another agency once that assessment has been approved by the Commission. EC 44320.2 (d) (3) further specifies that with respect to an alternative teaching performance assessment, the Commission shall: "Establish a review panel to examine each assessment developed by an institution or agency in relation to the standards set by the commission and advise the commission regarding approval of each assessment system."

A consortium of public and private institutions of higher education, represented by Stanford University, has developed an alternative teaching performance assessment and submitted this assessment to the Commission for review and approval. This assessment is the Performance Assessment for California Teachers, or PACT.

This agenda item provides (a) a description of the PACT assessment system; (b) a description of the review process of this assessment conducted by an expert review panel in relation to the Commission's adopted Assessment Design Standards; and (c) the expert panel's recommendation to the Commission for approval.

Background

Statutory Requirements for a Teaching Performance Assessment

Education Code Section 44320.2 provides for all of the following with respect to the qualities, role, and use of a teaching performance assessment:

- any locally developed performance assessment shall be based on assessment quality standards developed by the commission, which shall encourage the use of alternative assessment methods including portfolios of teaching artifacts and practices
- the performance assessment must ensure that oral proficiency in English is a criterion for scoring the performance of each candidate in the assessment
- the validity of assessment content and the reliability of assessment scores must be initially and periodically analyzed
- appropriate standards for satisfactory performance in the assessment must be established
- possible sources of bias in the performance assessment must be analyzed and action taken promptly to eliminate any bias that is discovered
- background information about candidates participating in the performance assessment must be collected and analyzed

- individual and aggregated results on the assessment must be reported and interpreted
- each assessment must be state approved and aligned to the California Standards for the Teaching Profession, and consistently applied to candidates in similar teacher preparation programs.
- to the maximum feasible extent, each performance assessment shall be ongoing and blended into the preparation program
- the performance assessment must be designed to provide formative assessment information during the preparation program for use by the candidate, instructors and supervisors for the purpose of improving the teaching knowledge, skill, and ability of the candidate
- the performance assessment results must be reported so that they may serve as one basis for a recommendation for the credential by the program sponsor
- the formative assessment information and the performance assessment results must be reported so that they may serve as one basis for the individual induction plan of the new teacher.

These requirements provide guidance to the developers of an alternative teaching performance assessment, are reflected in the Commission's Assessment Design Standards, and served as part of the orientation provided to the expert panel that reviewed the PACT submission (see Section II C below).

I. Description of the PACT Assessment System

The PACT assessment system represents the work of a consortium consisting initially of the following institutions:

- UC Berkeley
- UC Davis
- UC Irvine
- UC Los Angeles
- UC Riverside
- UC Santa Barbara
- UC Santa Cruz
- San Diego State University
- San Jose State University
- Mills College
- Stanford University

Over time, the consortium has expanded and now includes the following institutions:

UC	CSU	Private/Independent	District Intern
Berkeley	San Luis Obispo	Holy Names University	San Diego City Schools
Davis	Chico	Mills College	
Irvine	Channel Islands	Notre Dame de Namur	
Los Angeles	Dominguez Hills	Pepperdine	
Riverside	Monterey Bay	Stanford	
San Diego	Northridge	St. Mary's College	
Santa Barbara	Humboldt State	U. of the Pacific	
Santa Cruz	Sacramento State	U. of San Diego	
	San Diego State	USC	
	San Francisco State		
	San Jose State		
	Sonoma State		

PACT-related institutions credential 29.2% of newly credentialed teachers in California, based on the 2004-05 Title II report.

The design of the PACT assessment focuses on two assessment strategies: (1) the formative assessment of prospective teachers through “Embedded Signature Assessments” (ESAs) that occur throughout the teacher preparation sequence, and (2) the formative and summative assessment of prospective teachers through the “Teaching Event” that takes place during student teaching. The PACT scoring system is based on a series of scoring rubrics for the Teaching Event that are specifically developed for each of the content areas, and the scoring system includes formal training, calibration, and recalibration of assessors.

The ESAs represent course-embedded assignments that are considered to measure key competencies. The ESAs are completed by every student within a program, and are standardized across sections of a course within an institution. In addition, ESAs offer programs an opportunity to reflect their program philosophies and emphases in these assessments in a way that deepens candidate knowledge around the TPEs or goes beyond what is measured in the Teaching Event. ESAs are a product of lengthy deliberations among program faculty as to what is especially critical to becoming an effective teacher and how to best assess and measure these qualities. Within the PACT system, therefore, before candidates complete the Teaching Event, they have already received a great deal of ongoing support and formative feedback on the teaching competencies that are measured in the Teaching Event and, in some programs, on other dimensions of teaching as well.

The Teaching Event is both a formative and summative instrument. It was designed for use in making a summative decision about recommending a candidate for a Preliminary California Teaching Credential as well as to be diagnostic to support candidate growth. Specifically, the analytic rubrics were designed to provide formative feedback to the candidate to identify areas of strengths and weaknesses and to both promote candidate growth within the remainder of the program and to help lay a foundation for initial work in induction. Second, the Teaching Event

was designed to be implemented in a timeframe that allows for candidates to gain sufficient classroom teaching experiences prior to completion of the Teaching Event (usually during the second half of the program). However, this timeframe also provides them with sufficient time to receive feedback from course instructors on the quality of their performances and, if necessary, to remediate their performances. PACT programs currently sign a letter of agreement to systematically provide practice and feedback opportunities in each of the categories assessed. When the TPA requirement is effective in July 2008, All PACT member institutions will be required to sign a letter of agreement in which they agree to administer the Teaching Event (TE) in a timeframe that allows sufficient time for candidates to receive formative feedback (using the analytic rubrics) and receive remedial support prior to retaking the TE prior to graduation.

Further, when the TPA requirement becomes effective in July 2008, current PACT policy will be modified to require all Multiple Subjects teacher candidates in the PACT consortium to complete a Teaching Event in one content area (Literacy or Math), and three separate Teaching Event tasks for the other multiple subject content areas (Science and History/Social Science). Programs will have the option of having their candidates complete a Planning task, Instruction task, or Assessment task in each of these remaining content areas. The same rubrics that have been validated for use to score the Teaching Event (and have shown levels of scoring reliability that meet professional standards) will be used to score these individual tasks. In this way, all aspects of TPE 1 will be covered by PACT.

A more detailed description of the Teaching Event provided by the PACT developers follows:

The Teaching Events are subject-specific assessments linked to the California content standards for students, and are integrated across four tasks: planning, instruction, assessment and reflection (PIAR), with a focus on Academic Language embedded across the tasks. For each Teaching Event, candidates must plan and teach a learning segment comprised of 3 to 5 hours of instruction (i.e., an instructional unit or part of a unit), videotape and analyze their instruction, analyze student learning, and reflect on their practice. The Teaching Events are designed to measure and promote candidates' abilities to integrate their knowledge of content, students, and instructional context in making instructional decisions and reflecting on practice. By probing candidate thinking about student learning, completing the assessments provides important opportunities for mentoring and self-reflection. Additional benefits include focusing attention on the academic language development of all students, especially English learners and native speakers of varieties of English, as well as instructional strategies that are effective with a wide range of students.

In Task A (**"Planning Curriculum, Assessment, and Instruction"**), teacher candidates begin by describing the instructional context in which they will be teaching the learning segment for the Teaching Event. In order for raters to understand their teaching decisions, candidates are asked to write a commentary of about two pages that describes key characteristics of the class that affect the planning and teaching of the learning segment, such as characteristics of students in the class, the curriculum, and instructional context, including any constraints on their teaching. Candidates also complete an instructional context form in which they report the number of students in the class, the grade level of the class or any specialized

features, the number of special needs and English learners, the title of the textbook used (if any), and the number of available computers in the class and school. Candidates then provide an overview of their planned learning segment spanning 3-5 days of instruction, lesson plans for each lesson, assignments and other instructional materials for the learning segment.

In Task B (“**Implementing Instruction**”), candidates videotape one or more of their lessons from the learning segment, select up to 15-20 minute clips of the video (based on criteria set for each content area) and write a commentary on the unedited video clip(s) they have selected. In their commentary, candidates describe the context of the video clip (what happened before and after the clip); routines or working structures seen in the clip and how students were prepared for these routines; the ways in which the candidate engaged students with the lesson content; strategies used to address specific individual learning needs; and any language supports provided to students to understand the content or academic language.

In Task C (“**Assessing Student Learning**”), candidates collect and analyze student work from the learning segment. In the whole class learning commentary, candidates are asked to provide a context for the assessment, including a rationale for selection and the conditions under which students completed it; summarize student learning across the whole class relative to the learning goals; and discuss what most students seem to have understood and any misunderstandings, confusions, or special needs. In addition, candidates propose next steps in instruction based on their analysis of student learning. In the individual student learning commentary, candidates select two students in the class (who represent different instructional challenges) to focus on in analyzing student learning over time. In this task, candidates collect and analyze three samples of each student’s work that reflect his or her growth or progress with respect to a central goal of the class. Candidates are also asked to describe the feedback provided to students on their work.

In Task D (“**Analyzing and Reflecting on Teaching and Learning**”), candidates are prompted to reflect daily on their lessons after each day of instruction. At the end of the learning segment, candidates are asked to reflect on what they learned from their teaching of the learning segment and to describe what they would do differently if they were to teach the same content to the same group of students. They are also prompted to explain how their proposed changes would improve the learning of their students and to cite specific evidence and theoretical perspectives and principles that inform their analyses.

Although Academic Language is not a task in the Teaching Event, it comprises an analytic category in the scoring rubrics. The Academic Language rubric is scored based on evidence drawn from all of the tasks. Teacher candidates are prompted in the Planning and Instruction tasks to describe how their lessons and instruction help to build students’ acquisition and development of Academic Language. For example, in Task A, candidates are prompted to describe the language demands of the learning and assessment tasks that are likely to be challenging for their students. They are also

asked to describe how they planned to support students in meeting those language demands. Task B asks candidates to describe any language supports they used to help students understand the content and/or academic language. Task C asks candidates to discuss the progress in learning over time for two students, one of whom must be an English learner or another student who is struggling with academic English. Reflection on the successes and problems in each lesson with respect to developing language proficiency is prompted in Task D.

The PACT Scoring System

Each local campus will have a group of subject-specific trainers or will enter into agreements with other programs in the PACT consortium to share one or more trainers. The trainers will be prepared through a rigorous Training of Trainers program that will be repeated annually. Trainers will need to reach a calibration standard in order to be eligible to work as a trainer. The trainers will then assume a set of responsibilities that include training, calibrating, and supervising local scorers.

All Teaching Events will be independently scored at least once by trained and calibrated scorers at each local campus. All Teaching Events with scores that do not meet the established passing standard and borderline scores (those just above the passing standard) will be scored by one additional scorer, another trained local scorer, and the evidence read by the chief trainer, to ensure the reliability of the scores. In addition, as Teaching Events are scored, a randomly selected stratified sample of 10% of TEs from across the score levels and across scorers will be double-scored. If the scores given by two different scorers conflict by two or more rubric levels on any rubrics, or the differences in scores result in different pass/fail outcomes, the trainer in the specific content area will also score the Teaching Event to resolve discrepancies. Trainers will monitor the double-scoring by examining the scores for Teaching Events that were double-scored and conducting “read-behinds” for scores that are widely discrepant. The chief trainer will identify scorers who are drifting and will work with them to again achieve calibration by discussing the discrepant scores and helping the scorers to understand the differences between levels on rubrics that appear to be problematic for the scorers.

If there is sufficient evidence that local campuses continue to have large discrepancies with audit scores in the second year, external trainers will be sent to conduct the local campus training and supervise scoring.

Every third year, a central standardized scoring model will be used to provide another check on the consistency of training and the scoring process and the reliability and validity of scores. Under this model, local scorers will be convened at central scoring sites within a region to be trained and calibrated, and to score teaching events.

The PACT Passing Score Standard

The original passing score standard for the PACT assessment was that candidates pass the Teaching Event if they pass all five rubric categories (Planning, Instruction, Assessment, Reflection, and Academic Language) and have no more than 3 scores of “1” across the tasks. The cut score for each category was as follows (1.66 for Planning – (1 out of 3 scores can be a

1); (1.5 in Instruction, Assessment, Reflection, and Academic Language (1 out of 2 scores can be a 1).

Following feedback from the expert review panel (see below) concerning the passing score standard, in order to fully satisfy the Assessment Design Standards, the PACT developer convened a panel on August 16, 2007 of in-service teachers (teaching at both the elementary and secondary levels) and BTSA support providers who work with new teachers. These teachers and support providers were introduced to the Teaching Event prompts, rubrics, and scoring scales, and followed the same processes used in standard setting with other PACT standard setting panels. They were asked to independently arrive at a passing standard, with the underlying reasoning provided through discussions within a consensus decision-making process. The teacher panel recommended that the current passing standard be amended to allow *no more than two scores at Level 1 across the five categories*. Their amendment was submitted to all members of the original standards setting policy panel and was ratified by the panel.

The Commission's Assessment Design Standards (Standard 19 (i)) require that the passing score standard on any alternative teaching performance assessment must be equivalent to or more rigorous than the passing score standard on the Commission's teaching performance assessment. The PACT developer accurately notes that any process used to determine equivalence in the rigor of the two passing standards must take into account design differences between the two TPA systems. In the absence of a formal empirical study of equivalence the PACT developer has provided a clarification regarding the conceptual basis of the PACT passing score standard: "When we designed the Teaching Event rubrics, we purposefully chose to develop analytic rubrics in order to provide differentiated information on candidate performance to candidates, their supervisors, and their BTSA support providers. We also chose to construct 2-3 rubrics for every Teaching Event category (Planning, Instruction, Assessment, Reflection, and Academic Language - PIARA) to represent a small number of independent dimensions for each category. The Standard Setting panels felt that to pass the PACT Teaching Event, candidates should show a reasonable level of competence in every category. This is reflected in the final passing standard, which requires that candidates meet the cut score on every category of the Teaching Event. Candidates must pass all five PIARA categories, just as the CalTPA requires that all four tasks are passed."

II. The Review Process

A. Overview

The initial PACT submission was received by the Commission in March 2007, with an addition to this submission received on June 21, 2007 following conversations of the PACT sponsor with Commission staff. The review process included both of these documents as constituting the official PACT assessment system submission.

The review and approval process for the PACT alternative teaching performance assessment followed the Commission's standard process for reviewing program submissions. Applications from qualified educational assessment experts were invited and reviewed, and an expert review panel was appointed by the Executive Director to conduct the review and make a recommendation to the Commission. The purpose of the review process was for the panel to determine if the responses to the Commission's standards met each standard or if the program

sponsor needed to provide additional information before a standard was deemed to have been met. This is a positive review process designed to help a program sponsor be successful in the process if at all possible.

B. The Expert Review Panel

The expert review panel for alternative teaching performance assessments was comprised of six experienced measurement and assessment specialists. Three panel members were directors/administrators of districtwide testing and assessment in K-12 public school districts; two were faculty members in the area of measurement and assessment at institutions of higher education; and one member was a testing expert from the California Department of Education.

C. The Expert Panel's Review Process and Timeline

The review panel met in Sacramento on June 25, 2007. The Commission's standard panel review process was implemented for the review of the alternative teaching performance assessment. At the meeting, Commission staff discussed with the group the statutory requirements relating to the qualities, role and use of the teaching performance assessment and provided an orientation to each of the Commission's two Assessment Design Standards and their individual elements. The Assessment Design Standards are provided for reference as Attachment A. Staff also explained that the intent of the review process was not only to examine the submitted materials against each element of each of the Commission's Assessment Design Standards, but also to provide helpful feedback to the assessment sponsor if the panel found that a particular standard needed more information before a determination could be made that the standard was met. Staff explained to the panel that the overall orientation and goal of the process was to help the assessment sponsor ultimately get to "yes" in terms of the recommendation of the panel if this were at all possible.

The members of the expert review panel took their responsibilities seriously. Each panel member had thoroughly read the materials submitted by the PACT consortium prior to the panel meeting, and all engaged in highly professional conversations on measurement issues relating to performance assessment during the meeting. During the meeting, each response to the standards and their elements provided by the PACT sponsor was discussed and consensus was reached as to whether the information provided met the standards. As part of its deliberative process, the expert panel also reviewed the data provided by the PACT developer concerning content validity and scoring reliability of the PACT assessment system, as required by statute.

The review panel determined at the June 25, 2007 meeting that additional information was needed with respect to some of the standards, and feedback was provided in written form on July 6, 2007 to the PACT consortium sponsor, Stanford University. Commission staff had taken notes during the panel meeting, had summarized this information into the Commission's standard program feedback format, and had received confirmation from each of the panel members that the written summary accurately represented the panel's feedback before the feedback document was sent to the PACT sponsor.

Following several email and phone contacts with the PACT developers, Commission staff held a meeting with Ray Pecheone and Kendyll Stansbury, representing the PACT consortium, on

August 22, 2007 at the Commission's office to review the panel's feedback and discuss the areas where the panel had indicated that additional information was needed. The orientation at that meeting was to try to help the PACT developers respond to the panel's feedback in a manner that would be responsive to the standards as well as to the additional information and/or clarifications requested by the panel. The PACT representatives indicated they felt the review panel members had taken their roles seriously, and were impressed at the thoroughness with which the review panel had addressed the information provided in the initial PACT submission.

Additional information from the PACT developers in response to the review panel's feedback was received on September 6, 2007. The information was sent out electronically to the members of the expert review panel on the same day. Each panel member was asked to determine individually if, based on the additional information, the specific standards had been met. Each panel member has indicated that the complete PACT submission, including both the original submission and the additional information provided in response to the review panel's feedback, now meets the Commission's Assessment Design Standards.

C. The Expert Panel's Recommendation to the Commission

The expert panel has determined that the PACT assessment system (including the original design submitted in March and June 2007 and the modifications to the design indicated within the additional information submitted in September 2007) meets the Commission's Assessment Design Standards. Therefore, the expert panel recommends to the Commission that the Performance Assessment for California Teachers, or PACT, be approved as provided for under Education Code Section 44320.2 (d)(3).

ATTACHMENT A

ASSESSMENT DESIGN STANDARDS

Assessment Design Standard 1: Assessment Designed for Validity and Fairness

(Assessment Design Standard 1 Applies to Programs that Request Approval of Alternative Assessments)

The sponsor of the professional teacher preparation program requests approval of a Teaching Performance Assessment (TPA) in which complex pedagogical assessment tasks and multi-level scoring scales are linked to the Teaching Performance Expectations (TPEs) in Appendix A. The program sponsor clearly states the intended uses of the assessment, anticipates its potential misuses, and ensures that local uses are consistent with the statement of intent. The sponsor maximizes the fairness of assessment design for all groups of candidates in the program, and ensures that the established passing standard on the TPA is equivalent to or more rigorous than the recommended state passing standard.

Required Elements for Assessment Design Standard 1: Assessment Designed for Validity and Fairness

- 1(a) The Teaching Performance Assessment includes complex pedagogical assessment tasks to prompt aspects of candidate performance that measure the TPEs. Each task is substantively related to two or more major domains of the TPEs. For use in judging candidate-generated responses to each pedagogical task, the assessment also includes multi-level scoring scales that are clearly related to the same TPEs that the task measures. Each task and its associated scales measure two or more TPEs. Collectively, the tasks and scales in the assessment address key aspects of the six major domains of the TPEs. The sponsor of the professional teacher preparation program documents the relationships between TPEs, tasks and scales.
- 1(b) To preserve the validity and fairness of the assessment over time, the sponsor may need to develop and field-test new pedagogical assessment tasks and multi-level scoring scales to replace or strengthen prior ones. Initially and periodically, the sponsor analyzes the assessment tasks and scoring scales to ensure that they yield important evidence that represents candidate knowledge and skill related to the TPEs, and serves as a basis for determining entry-level pedagogical competence to teach the curriculum and student population of California's K-12 public schools. The sponsor records the basis and results of each analysis, and modifies the tasks and scales as needed.
- 1(c) Consistent with the language of the TPEs, the sponsor defines scoring scales so different candidates for credentials can earn acceptable scores on the Teaching Performance Assessment with the use of different pedagogical practices that support implementation of the K-12 content standards and curriculum frameworks. The sponsor takes steps to plan and anticipate the appropriate scoring of candidates who use pedagogical practices that are educationally effective but not explicitly anticipated in the scoring scales.

- 1(d) The sponsor develops scoring scales and assessor training procedures that focus primarily on teaching performance and that minimize the effects of candidate factors that are not clearly related to pedagogical competence, which may include (depending on the circumstances) factors such as personal attire, appearance, demeanor, speech patterns and accents that are not likely to affect student learning.
- 1(e) The sponsor publishes a clear statement of the intended uses of the assessment. The statement demonstrates the sponsor's clear understanding of the implications of the assessment for candidates, the public schools, and K-12 students. The statement includes appropriate cautions about additional or alternative uses for which the assessment is not valid. Before releasing information about the assessment design to another organization, the sponsor informs the organization that the assessment is valid only for determining the pedagogical competence of candidates for initial teaching credentials in California. All elements of assessment design and development are consistent with the intended use of the assessment for determining the pedagogical competence of candidates for Preliminary Teaching Credentials in California.
- 1(f) The sponsor completes content review and editing procedures to ensure that pedagogical assessment tasks and directions to candidates are culturally and linguistically sensitive, fair and appropriate for candidates from diverse backgrounds. The sponsor ensures that groups of candidates interpret the pedagogical tasks and the assessment directions as intended by the designers, and that assessment results are consistently reliable for each major group of candidates.
- 1(g) The sponsor completes basic psychometric analyses to identify pedagogical assessment tasks and/or scoring scales that show differential effects in relation to candidates' race, ethnicity, language, gender or disability. When group pass-rate differences are found, the sponsor investigates to determine whether the differences are attributable to (a) inadequate representation of the TPEs in the pedagogical tasks and/or scoring scales, or (b) overrepresentation of irrelevant skills, knowledge or abilities in the tasks/scales. The sponsor acts promptly to maximize the fairness of the assessment for all groups of candidates and documents the analysis process, findings, and action taken.
- 1(h) In designing assessment administration procedures, the sponsor includes administrative accommodations that preserve assessment validity while addressing issues of access for candidates with disabilities.
- 1(i) In the course of developing or adopting a passing standard that is demonstrably equivalent to or more rigorous than the State recommended standard, the sponsor secures and reflects on the considered judgments of teachers, the supervisors of teachers, the support providers of new teachers, and other preparers of teachers regarding necessary and acceptable levels of proficiency on the part of entry-level teachers. The sponsor periodically re-considers the reasonableness of the scoring scales and established passing standard.

Assessment Design Standard 2: Assessment Designed for Reliability and Fairness

(Assessment Design Standard 2 Applies to Programs that Request Approval of Alternative Assessments)

The sponsor of the professional teacher preparation program requests approval of an assessment that will yield, in relation to the key aspects of the major domains of the TPEs, enough collective evidence of each candidate's pedagogical performance to serve as an adequate basis to judge the candidate's general pedagogical competence for a Preliminary Teaching Credential. The sponsor carefully monitors assessment development to ensure consistency with the stated purpose of the assessment. The Teaching Performance Assessment includes a comprehensive program to train and re-train assessors. The sponsor periodically evaluates assessment design to ensure equitable treatment of candidates. The assessment design and its implementation contribute to local and statewide consistency in the assessment of teaching competence.

Required Elements for Assessment Design Standard 2: Assessment Designed for Reliability and Fairness

- 2(a) In relation to the key aspects of the major domains of the TPEs, the pedagogical assessment tasks and the associated directions to candidates are designed to yield enough evidence for an overall judgment of each candidate's pedagogical qualifications for a Preliminary Teaching Credential. The program sponsor will document sufficiency of candidate performance evidence through thorough field-testing of pedagogical tasks, scoring scales, and directions to candidates.
- 2(b) Pedagogical assessment tasks and scoring scales are extensively field-tested in practice before being used operationally in the Teaching Performance Assessment. The sponsor of the program evaluates the field-test results thoroughly and documents the field-test design, participation, methods, results and interpretation.
- 2(c) The Teaching Performance Assessment system includes a comprehensive program to train assessors who will score candidate responses to the pedagogical assessment tasks. An assessor training pilot program demonstrates convincingly that prospective and continuing assessors gain a deep understanding of the TPEs, the pedagogical assessment tasks and the multi-level scoring scales. The training program includes task-based scoring trials in which an assessment trainer evaluates and certifies each assessor's scoring accuracy in relation to the scoring scales associated with the task. When new pedagogical tasks and scoring scales are incorporated into the assessment, the sponsor provides additional training to the assessors, as needed.
- 2(d) In conjunction with the provisions of Standard 22, the sponsor plans and implements periodic evaluations of the assessor training program, which include systematic feedback from assessors and assessment trainers, and which lead to substantive improvements in the training as needed.

- 2(e) The program sponsor requests approval of a detailed plan for the scoring of selected assessment tasks by two trained assessors for the purpose of evaluating the reliability of scorers during field-testing and operational administration of the assessment. The subsequent assignment of one or two assessors to each assessment task is based on a cautious interpretation of the ongoing evaluation findings.
- 2(f) The sponsor carefully plans successive administrations of the assessment to ensure consistency in elements that contribute to the reliability of scores and the accurate determination of each candidate's passing status, including consistency in the difficulty of pedagogical assessment tasks, levels of teaching proficiency that are reflected in the multilevel scoring scales, and the overall level of performance required by the Commission's recommended passing standard on the assessment.
- 2(g) The sponsor ensures equivalent scoring across successive administrations of the assessment and between the Commission's model and local assessments by: using marker performances to facilitate the training of first-time assessors and the further training of continuing assessors; monitoring and recalibrating local scoring through third party reviews of scores that have been assigned to candidate responses; and periodically studying proficiency levels reflected in the adopted passing standard.
- 2(h) The sponsor investigates and documents the consistency of scores among and across assessors and across successive administrations of the assessment, with particular focus on the reliability of scores at and near the adopted passing standard. To ensure that the overall construct being assessed is cohesive, the sponsor demonstrates that scores on each pedagogical task are sufficiently correlated with overall scores on the remaining tasks in the assessment. The sponsor demonstrates that the assessment procedures, taken as a whole, maximize the accurate determination of each candidate's overall pass-fail status on the assessment.
- 2(i) The sponsor's assessment design includes an appeal procedure for candidates who do not pass the assessment, including an equitable process for rescoring of evidence already submitted by an appellant candidate in the program.

